Weekly Report

2018.1126-2018.1202

1.This Week

Deep Learning Power Grid Project:
1.We implemented a simple system that we can interact with to observer the patterns of the clustering results and we make several interesting findings:

- Nearly half of the dataset is clustered into a group that contains only one data point. The grey dots in the picture below indicates clusters (referred to as **individual clusters**) with only one data instance; the black dots are clustered into groups of multiple dots (referred to as **common clusters** below).

- Most of the common clusters contains outliers that is far away from the center of the cluster, as shown in the pictures below. After in-depth analyzing, we notice that this is because: in our data instances, we have two fault center in each instance. As for the outliers, the two fault centers of the outlier are all very close to the other instances in the cluster (fault centers in theses instances) when comparing by the numerical value of electrical distance (ELD). But when we check the ranking of the ELD, one of the fault center doesn't have a high ranking

(which means there are a lot of buses that are closer to the fault center in other data instances.).

- **Individual clusters** have a similar numerical value of ELD to nearby **common clusters**. But similarly, when we compute the distance between fault_center _of_individual_cluster to fault_center_of_nearby_common_clusters, we find other buses (not fault center) are more close to the fault_center_of_nearby_common_clusters. (The ELD ranking is low.)

2.Prof. Deng suggests that we should try to compute the correlations between buses so that we may find something. Actually, we have tried transfer entropy to dig the correlation but it is hard to explain. We now plan to try more simple correlation metrics such as Pearson correlation coefficient.

Power flow Project

1.The progress of this project：

- Most coordinated interactions are finished.
- A simple demo version is finished.
- We change the style of the interface twice to improve the space utilization.
- We plan to record a demo video when we finished the style revision on Monday.

2.The first two sections of the technical report is finished. There is only one section left: the one that introduces out system. I plan to finish it the next week.

Working Hour: (except nap and eat time)
8-9 hours / week day (sick on Monday)
8 hours on Sunday
Total Working Hour this week: 42 hours.

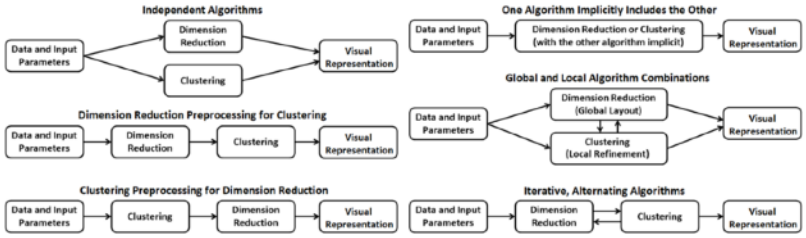
Other

1.Talk with Prof. Deng to learn his ideas and suggestions.
2.Learn about basic reinforcement learning ideas.

Paper Reading

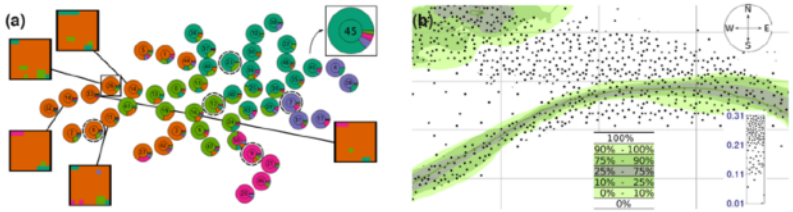
1.Towards a Systematic Combination of Dimension Reduction and Clustering in Visual Analytics

This paper talks about connections and differences between dimension reduction and clustering techniques. Most importantly, it presents six ways (pipelines) to combine these two techniques together, as shown in the picture below. This paper offers me a new way to look into these two techniques, especially dimension reduction. It is more than a tool to present high dimensional data. Dimension reduction and clustering can actually work together to support, revise, alternate each other.



2.Visualizing Confidence in Cluster-based Ensemble Weather Forecast Analyses

This paper presents a visual analytics solution for analyzing the sensitivity of clustering results of ensemble weather forecast data when changes take place in a selected region. It demonstrates the main problem that whether a cluster is representative and what changes make a cluster unstable. This problem is also the most basic one for all clustering methods but can hardly be solved by pure numerical solutions.



3.Temporal Sequence Learning and Data Reduction for Anomaly Detection

This paper presents an instance-based learning (IBL) technique for temporal sequence anomaly detection. The basic idea is to transform temporal sequences into a metric space via a chosen similarity measure that encodes intra-attribute dependencies. Instance-selection methods and clustering are used to reduce data storage requirements.

4.A Markov Chain Model of Temporal Behavior for Anomaly Detection

This paper presents an anomaly detection technique to detect intrusions into computer and network systems. It uses a Markov chain model to represent a temporal profile of normal behavior in a computer and network system. The Markov chain model is learned from historic data and the observed behavior of the system is analyzed to infer the probability in the model. A low probability indicates an anomalous behavior.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	12.15	1.Have interesting findings for the dataset.
SQC Paper	-	1.Delayed
Power Flow Project	December	1.Finish the system demo of the project. 2.The technical report are nearly finished(except the section that introduces the system prototype).